DOCUMENT RESUME

ED 287 289 FL 016 945

AUTHOR Hamp-Lyons, Liz
TITLE Performance Profiles for Academic Writing.
PUB DATE 87
NOTE 16p.; In: Language Testing Research; Selected Papers from the Colloquium (Monterey, California, February 27-28, 1986); see FL 016 938.
PUB TYPE Reports - Evaluative/Feasibility (142) -- Speeches/Conference Papers (150)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS English (Second Language); *English for Academic Purposes; Graduate Students; Higher Education; *Interrater Reliability; *Language Tests; Profiles; *Scoring; Second Language Instruction; Standardized Tests; Test Validity; *Writing (Composition); *Writing Evaluation; Writing Skills
IDENTIFIERS *British Council

ABSTRACT

A study examined the scoring procedure for the second part of the modular section of the English Language Testing Service (ELTS) academic writing test. The scoring is done by external raters according to procedures and a scale specified for the test, resulting in a performance profile. The report chronicles the development of the procedures and criteria, and examines the validity of the assessment technique as applied to the first item and the interrater reliability of a group of inexperienced raters. Problems and potential of this type of assessment are discussed. Appendices contain the original and revised assessment scales and the profile grid used in reporting scores. (MSE)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

PERFORMANCE PROFILES FOR ACADEMIC WRITING

Liz Hamp-Lyons
University of Michigan

ED287289

One of the many innovative features of the British Council's English Language Testing Service (ELTS), introduced in 1980, was the inclusion of a direct test of writing. The context of the test is the testing of the English proficiency of overseas non-native English speaking, mainly postgraduate, students who are applying for scholarships to British universities and other tertiary education institutions, and who are normally applying for scholarships from the British Council or one of the agencies whose funds are administered by the British Council.

Although direct testing of writing was very common until the 1930's or 40's, and was indeed the only test method in the 1800's, in the structuralist-psychometric era 'essay tests' had fallen into disrepute and disfavor as unreliable. The emphasis on language as communication in the early 1970's and the humanistic trends of the late 1970's, however, were reflected in an emphasis on test validity, and led to a search for tests which would combine the essential qualities of reliability with validity. In addition, developments in ESP emphasized face validity among the other validities, and led to a particular interest in performance testing. The British Council's decision to include a direct writing test in the ELTS battery was, then, a logical part of a general pattern in language teaching and testing.

The ELTS writing test is the second part of the Modular section of the ELTS (hence the abbreviation 'M2'); there are six Modules, and the candidate takes whichever is most closely akin to her/his own field of specialization. I do not propose in this chapter to discuss the specific versus general issues at all, focusing instead on one part of M2 which is common across Modules, that is, the scoring procedure.

M2 consists of two compulsory questions, each based on an input text which the candidate has previously read in another part of the test. The test lasts 40 minutes, and the first question (recommended time 25 minutes) has come to be described as "divergent," in that it requires the candidate to consider the information in the input text in relation to the question, but also expects the candidate to bring a personal response to the question, for example by relating it to her/his own country or own special subject interest. The second question (recommended time 15 minutes) has come to be known as "convergent," in that it requires the candidate to stay very close to the input text, extracting and organizing the information to fit the needs of the question. (Questions are confidential and may not be divulged.) I will focus here on the development of the scoring procedures for the first question.

When M2 was first introduced, scoring was done with the aid of a short paragraph explaining the need to value communicative quality more than structural and surface features, but the main guide for the rater was a set of performance descriptions each associated with a performance level from 1-9, coupled with an example of performance at each level (i.e., a benchmark paper).

FL016945

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

R. Schoffman

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

X This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

These nine levels and associated descriptions have become known as 'band descriptors', and I will refer to them throughout as the 'Original Assessment Scale', and to this scoring method as the 'Original Method.' The Original Assessment Scale is given in Appendix A.

It became clear that raters, most of whom were British Council ELT officers working in centers outside Britain and often in isolation from other M2 raters, needed firmer guidance in rating M2. An M2 Assessment Guide was written, piloted, revised and put into operation in 1985. This Assessment Guide took the criteria which had been implicit in the original general explanation of what should be valued in M2 answers and made these explicit. Each criterion was extensively characterized and some key problems raised by raters (e.g., "How long must an answer be before it can be looked at as 'communication'?" or "What constitutes plagiarism and how is it to be rated?") were tackled. The Guide took a self-study standardization approach, and included a set of 'criterion' papers for trial scoring, with discussion of how the standardization team handled them, and a further set for refresher scoring. The criteria for assessment of the first question presented in the 1985 Assessment Guide are:

Communicative quality
Organization
Argumentation
Linguistic accuracy and appropriacy

The rater is required to skim-read the essay three times: the first time, the rater focuses on communicative quality, i.e., a holistic reading, and makes a broad judgment which encompasses a three-band range (e.g., 2-5, 7-9, etc.). The second time the rater focuses on organization and argumentation and narrows the original judgment to a two-band range (2-3, 7-8, etc.). Finally, the rater reads again focusing on linguistic accuracy and appropriacy and decides on a single band from the two-band range, which is the final score for this question. This procedure is know as the 'Global Method.'[1]

But the Guide did not revise the original bands, and this presents several difficulties for the rater. First, the labels 'Competent Writer,' 'Marginal Writer,' etc., are difficult to interpret: in a trial with 20 raters I found that only 14 were able to correctly match the labels with the descriptions and number the sets in the correct order 1-9. My experience in training raters also suggests that the labels tend to discourage raters from looking closely at the full performance descriptions. But a more serious difficulty, and this is a recurring problem in rating essay tests, is that a single scale implies a unidimensional view of writing proficiency, and necessitates the treatment of each essay as existing at a single level. This would not be problematic if all writers did in fact write at a single performance level (thus manifest a 'flat' profile), but reports from raters indicated that sometimes a rater had problems rating an essay because she or he could not see one uniform level in the essay. Detailed observation of how raters

[1] A copy of the relevant section of the M2 Assessment Guide may be obtained by writing to the Consultant, ELTS; ELLD, the British Council; 10, Spring Gardens, London, SW1A 2BN, England.

actually rate revealed that the raters' instincts were accurate: the essays they had problems with were assessable using the criteria which had been developed, but only by looking at each criterion separately. The writers of such essays showed greater proficiency on some criteria than others. In ELTS terminology this multidimensional proficiency is referred to as a 'marked' profile. While the term was introduced to describe variations in proficiency across skills, it applies equally across the dimensions of a single skill such as writing.

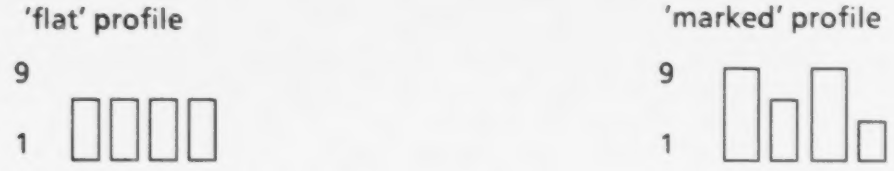

Figure 1: 'Flat' versus 'marked' profiles.

The Assessment Guide, therefore, incorporated two scoring methods: the unidimensional assessment (Global Method) is applied first, to all essays, while the 'Profile Method' (referred to later in this paper as 'Profile Method 2') was developed for use with, and only with, problem essays. At the heart of this method is the Profile Grid which schematises each criterion separately and provides a scale with the numbers of the bands for each of the criteria (Figure 2). Raters are asked to circle a three-band range on each criterion on the Profile Grid, and can then either choose the mode as their final score, or total the mid-bands on the criterion and divide by 5.[2]

As stated above, the intention was that the Profile Method would only be used with problem essays, after an initial application of the unidimensional assessment (Global Method). It became clear, however, that some raters began to apply the Profile Method to every paper. This meant that instead of looking first at communicative quality and then moving through organization and argumentation to linguistic appropriacy and accuracy (i.e., macro to micro features) in the Global Method as had been intended, these raters began with linguistic features and moved in the opposite direction. This resulted in more emphasis being placed on linguistic features than had been intended by the test design. We may speculate as to the reasons for this preferred use of the Profile Method: it may be because there is no such thing (or, at least, that some raters perceive no such thing) as a 'flat' profile, that is a writer whose proficiency is the same on every aspect of the writing process; or it may be that the Profile Method artificially creates multiple samples, permitting an objectivisation of what is for some raters an uncomfortably subjective process.

---

[2] A copy of the relevant section of the M2 Assessment Guide may be obtained by writing to the Consultant, ELTS; ELLD, the British Council; 10 Spring Gardens London, SW1A, England.

Practical use of the Assessment Guide, then, indicated that it had improved matters considerably, but that there was room for further development. In particular, there were three key reasons for revisions of the original assessment scale. First, the criteria were not fully or consistently articulated in the original assessment scale: the test designers had themselves been searching for a sense of what the criteria were or should be. This could only be known as a consequence of the operationalization of the test. Revision would permit the scale to be brought into line with the Assessment Guide as a whole with a clear and consistent treatment of the same criteria. Detailed observations of raters during the development and piloting of the Guide had shown that raters found their task much easier when the performance descriptions were presented as direct linguistic parallels, with the same criteria in the same sequence in each. Second, revision would permit the careful and clear differentiation of the nine levels of performance on each of the criteria. Raters had reported that they found difficulty differentiating between bands 6 and 5 in particular on the original assessment scale. Clearly, the consistent format also helped in this area. Perhaps most importantly for the long term, revision would allow the integration of the profiling principle which is at the heart of the philosophy of the ELTS, by taking account of marked as well as flat profiles; a first attempt had been made in this direction in the Profile Method, but as noted above this system had weaknesses.

Since the criteria had stood the test of practical use well, the first stage of the revision, the construction of a new set of global performance descriptions to match these criteria, was not difficult, although it required more than one piloting to ensure that raters could satisfactorily distinguish between the levels all the way along the scale. One significant change implemented in the Revised Assessment Scale was the separation of linguistic accuracy and linguistic appropriacy into two criteria. This was done for two reasons: first, some raters had reported occasions on problem papers when they felt there was a difference in performance for the same student on features of accuracy compared to features of appropriacy; second, it was generally agreed that the linguistic qualities of essays should be given more emphasis than they were in the Original Assessment Scale. This separation had already been implemented in the Profile Grid, but was now made more explicit. When the Revised Assessment Scale was ready (Appendix C), I found in a trial with 20 raters that they were all able to correctly sequence the nine levels (bands) without access to labels. (Note that in the Revised Assessment Scale labels are not used.)

The second stage of the revision was to develop a Profile Scale to enable each criterion to be examined independently. This would enable the handling of problem papers in the same way as had been done with the Profile Grid, although at a finer level of detail. This simply involved separating out the criteria of the Revised Assessment Scale and presenting them in individual columns. These are referred to as the new Profile Scale and Profile Method 2 (PM2). (See Appendix D.)

The rater is asked to choose a single band to describe performance on each criterion, not the three-band range previously used. It was believed that when criteria are sufficiently precise there is no reason to work from imprecise ratings, and that the imprecision had added to the difficulties of score aggregating in the first Profile Method. In other ways, however, the new Profile Scale (Profile Method 2) presents the same problems of score aggregating as the first Profile Method did. There is no mathematical formula for score aggregating;

combining scores on organization and linguistic accuracy and calling the answer writing proficiency is much like adding two apples and three pears and calling the result a lemon. Nevertheless, it has to be done, since clearly those responsible for absolute acceptance/rejection decisions for university places or for scholarships must have a single number to use. Whatever ethics or aesthetics may desire, this is the practical reality. It must be the test developers' responsibility to advise the score consumers of their best estimate of the candidate's writing proficiency, treating as unidimensional that which experience has shown is not unidimensional. The way in which the separate scores are aggregated must reflect the belief of the test developers about what is important in writing performance for the specific context, and in what proportions compared to other dimensions entering the same equation. There is no single 'right answer.' The answer which was arrived at for the particular ELTS M2 context was to weight communicative quality twice and all the other criteria once. However, no one involved believes this is an insignificant decision, and it is one which may be revised in the future as a result of the study of the test in operation, which is always continuing.

This chapter has so far focused on the validity of the assessment of the first question of M2: let us now consider reliability. In a small study comparing the various scoring procedures developed so far for M2, 12 inexperienced raters worked in four teams of three, each team using one of the scoring procedures. The raters were chosen mainly for their availability and willingness, but also as being suitable candidates for positions in British Council DTO's (Direct Teaching Operations, i.e., British Council centres where English is taught), and therefore potential raters of M2 in the field. The four scoring procedures used were:

1. original assessment scale using the original bands (OM;Appendix A);
2. original assessment scale, combined with Profile Method 1 (PM1; Appendix B);
3. revised assessment scale and the global method (RM; Appendix C);
4. new Profile Scale and Profile Method 2 (PM2; Appendix D).

All the raters were given the same general introduction and a short training session, rating two answers by their assigned scoring procedure. Each team then rated the same ten answers, first giving an individual rating and then agreeing on a final rating. Investigation of the scores assigned by the raters as individuals as compared to the scores assigned by raters as teams showed that the original method (OM) resulted in the largest number of rater disagreements (defined as each rater having a different score, i.e. at least a three-band spread for the three scores): raters disagreed on five out of ten answers; one answer received ratings of 7, 5 and 3. The original assessment scale combined with Profile Method 1 (PM1) resulted in two cases of rater disagreement, the revised assessment scale (RM) resulted in one case of rater disagreement, and the revised Profile Method (PM2) resulted in no cases of rater disagreement. The average single rater reliabilities for the four methods were:

OM	.563
PM1	.864
RM	.883
PM2	.942

Using the Spearman-Brown prophecy formula the reliabilities with three raters are estimated thus:

OM	.790
PM1	.950
RM	.960
PM2	.997

(It can be seen that with a more reliable method there is proportionately less additional reliability for more raters.)

On this preliminary check, then, the development of the new rating scale seems to have been of marked benefit to reliability. The use of a performance profile approach in the form of both the profile grid (PM1) and the profile scale (PM2) also contributes something to reliability: in the case of the addition of the first profile method to the original assessment, the result is a major increase in reliability; in the case of the addition of the second profile method to the revised assessment scale, the increase in reliability is only slight, and the single rater reliability for RM is, on this sample, more than adequately reliable.

In the British Council context, as explained above, the practical reality is that M2 is scored by a single rater, often working in considerable isolation from other raters. What must interest us in this context is a high single-rater reliability rather than any theoretically but not operationally achievable multiple-rater reliability. For this purpose any of the methods except the original one is acceptable.

The correlations between the four methods were generally quite high. Listed below are the correlations between the aggregate scores for sets of logical 'pairs:'

OM/PM1	.908
RM/PM2	.920
OM/RM	.845
PM1/PM2	.929

The other two correlations are .827 for OM with PM2, and .864 for PM1 with RM.

It can be seen that the highest correlation is for PM1 with PM2. These two methods are very similar in allowing the rater to treat each essay as a multiple sample: conceptually they share a view of writing as (at least potentially) multidimensional. We may hypothesize that the profile grid, although it was without any descriptions for the different criteria at each level, achieved what had been intended simply by allowing the rater the 'space' in which to respond. PM2 takes this much further than PM1, but it may be more an administrative convenience than anything else, since the descriptors are already present in the global version of the revised assessment scale: all the profile version of the scale does is break them up conveniently.

The high correlation for RM/PM2 is an important confirmation that these two related methods are yielding comparable scores, which is essential when two methods are used as alternate possibilities with the same set of candidates, as these two methods are. The correlation for OM/PM1, while not quite as high, is similarly at a reassuring level. We do not, of course, yet have data to show

whether similar correlations will be achieved in the field, with real 'problem' answers.

It is worth noticing that the choice of scoring procedure appeared to have a slight but noticeable influence on the resulting score level: Table 1 shows that OM tends to be more generous than the other three methods, and that PM2 tends to be more stringent. It would appear that as the scoring method has been refined and become more rigorous, it has also become more stringent. When averaged, these differences are quite small, but on a single-scorer test any differences may be very dramatic for any one individual. It is, therefore, heartening that the trend in the development of the methods has been towards increasing reliability. Table 1 shows the aggregate score for each answer for each scoring method:

Table 1

Aggregate Scores: Essay x Scoring Method

	Method			
	OM	**PM1**	**RM**	**PM2**
Essay				
1.	6*	7	7	7
2.	7	7	7	6*
3.	5	6*	5	5
4.	6	5	6	5
5.	8	8	7	7
6.	4*	3	2	2
7.	4	4	5	5
8.	4	3	4	3
9.	6	6	5	5
10.	5	5	4*	5

The single asterisk (*) indicates where there is an aggregate score which is noticeably different from the others: however, even in these cases the 'wild' score is only different by a single band (e.g., No. 1:7*:6:6:6:). The widest range of aggregate scores on the different scoring methods is three bands (e.g., No. 6:4*:3:2:). Nevertheless, this is a significant difference if it is, for example,the difference between a band 5 and a band 7: band 5 is unlikely to be considered acceptable for scholarship purposes without remedial English, while band 7 is almost certain to be acceptable as it stands. It can be noted in passing that anecdotal reports from score consumers, such as tutors on EAP courses, suggest that the M2 score has often been found to be over-generous. The increasing stringency noted above may provide a more accurate reflection of candidates' writing proficiency.

It must be remembered that the raters used for this investigation were not well (or uniformly) trained or experienced. However, the brief training they did receive was with this researcher, i.e., an experienced trainer/rater who thoroughly understood each of the methods and their rationale. We cannot know for sure what use is made of the Assessment Guide by raters in the field. Only after the new methods have been in use for some time and sufficient data have been collected and analyzed will we know whether similar patterns emerge.

Finally, it should be noted that the real advantage of the profile methods lies in their diagnostic function, especially in the case of PM2: if we can achieve not only more accurate information, but more information, we open tremendous potential for the use of test results in other contexts, and the test instrument makes gains in practicality. A testing system such as the ELTS is predicated on the belief that by administering tests of different skills, using different methods, and by reporting scores on each of these tests, not simply more but also better information is obtained about candidates, and as a result better decisions are made. If the decision made is for acceptance with some additional language programme, the test score information is available for diagnostic use. For a test such as M2, which is a direct performance test, and which according to experience reveals within-writer differences in some cases, extending the profiling to a more finely-tuned level not only aids reliability but is also a powerful tool for diagnosis and remediation. The potential of the instrument, as it now exists, for diagnosis and thence flexible placement is currently under investigation.

Appendix A

Original Assessment Scale ("OM")

BAND	BRIEF PERFORMANCE DESCRIPTION
9	**Expert Writer:** theme presented in a readable, intelligible, logical and interesting manner. Writes with complete accuracy and in the appropriate style. The reader is given a sense of mastery of the language and of the ability to handle the topic with complete competence.
8	**Very Good Writer:** theme presented clearly and logically, with accurate language forms and good style. Only very occasional inaccuracy or inappropriacy but which does not affect the communication. The reader can follow with no strain and will appreciate the argument expressed.
7	**Good Writer:** theme presented in a well-ordered, intelligible manner with well-structured and relevant supporting detail. Generally accurate in language and appropriate in style, but occasional lapses can affect the communication on first reading. The reader has, however, the impression of a functionally efficient writer.
6	**Competent Writer:** theme presented fairly logically and intelligibly. Reasonably accurate use of the language system. May have inaccuracies of style and presentation but showing an adequate functional competence. Can be read with only occasional strain put on comprehension.
5	**Modest Writer:** theme can be followed, but logical presentation may be broken and lack clarity or consistency. Several inaccuracies and style not always appropriate to presentation. May lack interest or variety, but the basic message is presented. The reader will have to strain on occasion to comprehend meaning.
4	**Marginal Writer:** theme can be followed with effort, and closer reading reveals lack of logical structure, clarity and consistency. Inaccurate vocabulary and sentence use coupled with inadequate connectors and cohesive features. Elements of information required may be omitted, repeated or inappropriately expressed. The reader has general difficulty in working out the message, though can eventually do so.
3	**Extremely Limited Writer:** elements of the information required are provided, but the presentation lacks any coherence. Uses over-simple sentence structure and impoverished vocabulary with continual errors and inappropriateness. Below level of functional competence though the reader may work out the general message.

2 **Intermittent Writer**: elements of the information required not provided, although a general meaning comes through intermittently. Either copies or produces strings of words. No real communication, although the reader may work out the general message.

1 **Unassessable Writer**: to be used for the true non-writer where no assessable strings of continuous English writing have been produced. OR: answer has been lifted 'en bloc' from Source Booklet, or a clearly irrelevant stock answer has been reproduced.

0 Should only be used where a candidate did not attend or attempt this part of the test in any way (i.e. did not submit an answer paper with his/her name and candidate number written on).

The British Council 1984

Appendix B

Profile Method ("PM1")

Profile Grid

Communicative Quality	9	8	7	6	5	4	3	2	1
Organization	9	8	7	6	5	4	3	2	1
Argumentation	9	8	7	6	5	4	3	2	1
Linguistic Appropriacy	9	8	7	6	5	4	3	2	1
Linguistic Accuracy	9	8	7	6	5	4	3	2	1

The British Council 1984

Appendix C

Revised Assessment Scale and the Global Method ("RM")

9 The writing displays an ability to communicate in a way which gives the reader full satisfaction. It displays a completely logical organizational structure which enables the message to be followed effortlessly. Relevant arguments are presented in an interesting way, with main ideas prominently and clearly stated, with completely effective supporting material; arguments are effectively related to the writer's experience or views. There are no errors of vocabulary, spelling, punctuation or grammar and the writing shows an ability to manipulate the linguistic systems with complete appropriacy.

8 The writing displays an ability to communicate without causing the reader any difficulties. It displays a logical organizational structure which enables the message to be followed easily. Relevant arguments are presented in an interesting way, with main ideas highlighted, effective supporting material and they are well related to the writer's own experience or views. There are no significant errors of vocabulary, spelling, punctuation or grammar and the writing reveals an ability to manipulate the linguistic systems appropriately.

7 The writing displays an ability to communicate with few difficulties for the reader. It displays good organizational structure which enables the message to be followed without much effort. Arguments are well presented with relevant supporting material and an attempt to relate them to the writer's experience or views. The reader is aware of but not troubled by occasional minor errors of vocabulary, spelling, punctuation or grammar, and/or some limitations to the writer's ability to manipulate the linguistic systems appropriately.

6 The writing displays an ability to communicate although there is occasional strain for the reader. It is organized well enough for the message to be followed throughout. Arguments are presented but it may be difficult for the reader to distinguish main ideas from supporting material; main ideas may not be supported; their relevance may be dubious; arguments may not be related to the writer's experience or views. The reader is aware of errors of vocabulary, spelling, punctuation or grammar, and/or limited ability to manipulate the linguistic systems appropriately, but these intrude only occasionally.

5 The writing displays an ability to communicate although there is often strain for the reader. It is organized well enough for the message to be followed most of the time. Arguments are presented but may lack relevance, clarity, consistency or support; they may not be related to the writer's experience or views. The reader is aware of errors of vocabulary, spelling, punctuation or grammar which

intrude frequently, and of limited ability to manipulate the linguistic systems appropriately.

4 The writing displays a limited ability to communicate which puts strain on the reader throughout. It lacks a clear organizational structure and the message is difficult to follow. Arguments are inadequately presented and supported; they may be irrelevant; if the writer's experience or views are presented their relevance may be difficult to see. The control of vocabulary, spelling, punctuation and grammar is inadequate, and the writer displays inability to manipulate the linguistic systems appropriately, causing severe strain for the reader.

3 The writing does not display an ability to communicate although meaning comes through spasmodically. The reader cannot find any organizational structure and cannot follow a message. Some elements of information are present but the reader is not provided with an argument, or the argument is mainly irrelevant. The reader is primarily aware of gross inadequacies of vocabulary, spelling, punctuation and grammar; the writer seems to have no sense of linguistic appropriacy, although there is evidence of sentence structure.

2 The writing displays no ability to communicate. No organizational structure or message is recognizable. A meaning comes through occasionally but it is not relevant. There is no evidence of control of vocabulary, spelling, punctuation or grammar, and no sense of linguistic appropriacy.

1 A true non-writer who has not produced any assessable strings of English writing. An answer which is wholly or almost wholly copied from the input text or task is in this category.

0 Should only be used where a candidate did not attend or attempt this part of the test in any way (i.e., did not submit an answer paper with his/her name and candidate number written on).

The British Council 1986

Appendix D

New Profile Scale and Profile Method 2 ("PM2")

	Communicative Quality	Organization	Argumentation	Linguistic Accuracy	Linguistic Appropriacy
9	The writing displays an ability to communicate in a way which gives the reader full satisfaction.	The writing displays a completely logical organizational structure which enables the message to be followed effortlessly.	Relevant arguments are presented in an interesting way, with main ideas prominently and clearly stated, with completely effective supporting material; arguments are effectively related to the writer's experience or views.	The reader sees no errors of vocabulary, spelling, punctuation or grammar.	There is an ability to manipulate the linguistic systems with complete appropriacy.
8	The writing displays an ability to communicate without causing the reader any difficulties.	The writing displays a logical organizational structure which enables the message to be followed easily.	Relevant arguments are presented in an interesting way, with main ideas highlighted, effective supporting material and they are well related to the writer's own experience or views.	The reader sees no significant errors of vocabulary, spelling, punctuation or grammar.	There is an ability to manipulate the linguistic systems appropriately.
7	The writing displays an ability to communicate with few difficulties for the reader.	The writing displays good organizational structure which enables the message to be followed without such effort.	Arguments are well presented with relevant supporting material and an attempt to relate them to the writer's experience or views.	The reader is aware of but not troubled by occasional minor errors of vocabulary, spelling, punctuation or grammar.	There are minor limitations to the ability to manipulate to linguistic systems appropriately which do not intrude on the reader.
6	The writing displays an ability to communicate although there is occasional strain for the reader.	The writing is organized well enough for the message to be followed throughout.	Arguments are presented but it may be difficult for the reader to distinguish main ideas from supporting material; main ideas may not be supported; their relevance may be dubious; arguments may not be related to the writer's experience or views.	The reader is aware of errors of vocabulary, spelling, punctuation or grammar, but these occasionally.	There is limited ability to manipulate the linguistic systems appropriately, but this intrudes only occasionally.

	Communicative Quality	Organization	Argumentation	Linguistic Accuracy	Linguistic Appropriacy
5	The writing displays an ability to communicate although there is often strain for the reader.	The writing is organized well enough for the message to be followed most of the time.	Arguments are presented but may lack relevance, clarity, consistency or support; they may not be related to the writer's experience or views.	The reader is aware of errors of vocabulary, spelling, punctuation or grammar which intrude frequently.	There is limited ability to manipulate the linguistic systems appropriately, which intrudes frequently.
4	The writing displays a limited ability to communicate which puts strain on the reader throughout.	The writing lacks a clear organizational structure and the message is difficult to follow.	Arguments are inadequately presented and supported; they may be irrelevant; if the writer's experience or views are presented their relevance may be difficult to see.	The reader finds the control of vocabulary, spelling, punctuation and grammar inadequate.	There is inability to manipulate the linguistic systems appropriately, which causes severe strain for the reader.
3	The writing does not display an ability to communicate although meaning comes through spasmodically.	The writing has no discernible organizational structure and a message cannot be followed.	Some elements of information are present but the reader is not provided with an argument, or the argument is mainly irrelevant.	The reader is primarily aware of gross inadequacies of vocabulary, spelling, punctuation and grammar.	There is little or no sense of linguistic appropriacy, although there is evidence of sentence structure.
2	The writing displays no ability to communicate.	No organizational structure or message is recognizable.	A meaning comes through occasionally but it is not relevant.	The reader sees no evidence of control of vocabulary, spelling, punctuation or grammar.	There is no sense of linguistic appropriacy.
1	A true non-writer who has not produced any assessable strings of English writing. An answer which is wholly or almost wholly copied from the input text or task is in this category.				